The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

CERTIFIED COPY OF PRIORITY DOCUMENT

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed M Cooke

Dated 13 February 2001

Patents Form 1/77

Patents Act 1977
(Rule 16)

THE PATENT OFFICE
E
22 DEC 2000
NEWPORT

23DEC00 E593892-1 D01463
P0177700 0.00-0031450

The Patent Office
Cardiff Road
Newport
South Wales
NP10 8QQ

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

1. Your reference — 30011097 GB

2. Patent application number *(The Patent Office will fill in this part)* — 22 Dec 2000 — 0031450.0

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

Hewlett-Packard Company
3000 Hanover Street
Palo Alto
CA 94304, USA

4965880 01

Patents ADP number *(if you know it)*

If the applicant is a corporate body, give the country/state of its incorporation — Delaware, USA

4. Title of the invention — Speech recognition

5. Name of your agent *(if you have one)*

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

Robert F Squibbs
Hewlett-Packard Ltd, IP Section
Filton Road
Stoke Gifford
Bristol BS34 8QZ

07563083001

Patents ADP number *(if you know it)*

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.
See note (d))

Yes

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	8
Claim(s)	2
Abstract	1
Drawing(s)	3 +3

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	1 ✓
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	fee sheet ✓

11. I/We request the grant of a patent on the basis of this application.

Signature R.F Squibbs Date Robert Francis Squibbs 21 December 2000

12. Name and daytime telephone number of person to contact in the United Kingdom K Nommeots-Nomm Tel: 0117-312-9947

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 08459 500505.

b) Write your answers in capital letters using black ink or you may type them.

c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you have answered 'Yes' Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) For details of the fee and ways to pay please contact the Patent Office.

Speech Recognition

Field of the Invention

The present invention relates to the automated recognition of speech.

Background of the Invention

The previous 40 years of spoken language systems work have been based on the belief that a computer could be trained in a human language and have a dialogue with a human. However after even all these years this remains a very hard problem and progress has been extremely slow. The resulting systems are not efficient and do not respond well to variations in accent, individual pronunciation, different ways of expressing a concept, noisy environments, variations in human language groups etc.

It is an object of the present invention to facilitate automated recognition of speech.

Summary of the Invention

The present invention is based on the creation of new spoken languages that can be easily understood by automated speech recognizers associated with equipment, the language being learnt by human users in order to speak to the equipment.

These spoken languages are hereinafter referred to as "Computer Pidgin Languages" or "CPLs", because like Pidgin languages in general, they are simplified in terms of vocabulary and structure. However, unlike normal human pidgin languages, the CPLs are languages specifically designed to minimize recognition errors by automated speech recognizers. In particular, a CPL language is made up of phonemes or other uttered elements that are not easily confused with each other by a speech recognizer, the uttered elements being preferably chosen from an existing language.

The present invention gives rise to CPL-enabled apparatus and methods of controlling equipment using CPLs. The invention also provides training systems for training human users to speak a CPL, and methods and systems for creating new CPLs.

More formally stated, according to one aspect of the present invention, there is provided an artificial spoken language for apparatus control, the language being made up of words formed from a set of phonemes, or other utterance elements, chosen for their low risk of confusion one with another by a speech recogniser.

According to another aspect of the present invention, there is provided apparatus with a voice-input interface including a speech recogniser adapted to recognise input in the form of an artificial language made up of words formed from a set of phonemes, or other utterance elements, chosen for their low risk of confusion one with another by the speech recogniser.

According to a further aspect of the present invention, there is provided a method of controlling apparatus wherein a user uses an artificial language to instruct the apparatus by voice input, the apparatus having an associated speech recogniser and said artificial language being made up of words formed from a set of phonemes, or other utterance elements, chosen for their low risk of confusion one with another by the speech recogniser.

According to a still further aspect of the present invention, there is provided a training system for teaching a person to use an artificial language, the system including means for receiving voice input from the user, means for detecting errors in the pronunciation and/or grammar of the voice input as compared to an expected input that conforms with said artificial language, and means for providing feedback to the user concerning the detected errors, said artificial language being made up of words formed from a set of phonemes, or other utterance elements, chosen for their low risk of confusion one with another by a speech recogniser.

According to a yet further aspect of the present invention, there is provided apparatus for use in constructing an artificial language, the apparatus comprising means for providing a set of phonemes or other utterance elements chosen for their low risk of confusion one with another by a speech recogniser, user input means for creating words by assembling together

said phonemes/utterance elements, and means for storing the set of created words as a new said artificial language.

Brief Description of the Drawings

Embodiments of the invention will now be described, by way of non-limiting example, with reference to the accompanying diagrammatic drawings, in which:

- **Figure 1** is a confusion matrix for a given set of phonemes;
- **Figure 2** is a diagram illustrating a system for creating a new CPL;
- **Figure 3** is a diagram illustrating speech recognizer adapted to interpret voice input according to any of multiple CPLs; and
- **Figure 4** is a diagram illustrating a training system for training a user to speak in a CPL.

Best Mode of Carrying Out the Invention

As already indicated, the present invention concerns the creation and use of spoken artificial languages (CPLs) that are adapted to be recognised by speech recognisers. A new CPL can be created as required, for example, for use with a new class of device.

A new CPL is preferably created by following the simple rules set out below:

1. Pick a subset of phonemes from a specific human language (such as English or Esperanto) that are not easily confused by an automated speech recognition, and are easily recognized. This subset may exhibit a dependency on the speech recognition technology being used; however, since there is generally a large overlap between the subsets of easily recognized phonemes established with different recognition technologies, it is generally possible to choose a subset of phonemes from this overlap area. It should also be noted that the chosen phoneme subset need not be made up of phonemes all coming from the same human language, this being done simply to make the subset familiar to a particular group of human users.
2. Make words up that are easily recognized and distinguished using the phonemes from the subset chosen in (1). The constructed words are preferably structured as CVC

(Consonant Vowel Consonant) like Japanese as this structure is believed to perform best in terms of recognition. Other word structures, such as "CV", are also possible.

3. Pick a filler sound that allows word boundaries to be easily distinguished (this step is optional, particularly where words are intended only to be used individually since silence then constitutes an effective filler).

4. Pick a simple grammar structure with very little ambiguity (again, this step is optional in the sense that where a CPL is based on single word commands, no grammar is required – other than that the command words are to be taken individually).

With respect to the choosing of a low-confusion-risk phoneme subset, a phone confusion matrix can be produced for a particular speech recognizer by comparing the input and output of the recognizer over a number of samples. This matrix indicates for each phone the degree of correlation with all the other phones. In other word, this matrix indicates the likelihood of a phone to be mistaken by others during the recognition process. Figure 1 shows an example confusion matrix produced from a British English corpus, (the illustrated matrix is produced without reference to a language model). This matrix is not symmetrical because the recognizer is not equally trained on all the phones (some of them being more frequent in the training set of data than others).

By examining the matrix, it is readily possible to ascertain which pairings of phonemes should be avoided if confusion is not to result.

Figure 2 illustrates a system 20 by which a user 2 can generate a new CPL . The system 20 is based on a computer running a CPL creation application 21 and storing in memory 22 the low-confusion-risk phoneme subset 23 for a language base (such as British English) selected by the user. This phoneme subset is presented to the user 2 (see arrow 25) who then uses the phonemes as building blocks for constructing new words which are stored back to memory (see arrows 26) as part of the new CPL 24. The user can also specify a grammar for the new CPL, this grammar being stored (see arrow 27) as part of the CPL. The system is also arranged to test out the chosen words for ease of recognition and lack of confusion on a target speech recognizer, the results of this test being fed back to the user;

this testing can either be done automatically (for example, whenever a new word is stored) or simply upon user request.

Whilst the human meaning associated with a CPL word is likely to be attributed at this stage (the CPL word may suggest this meaning in the base language), this is not essential. Furthermore, the mapping between the speech recognizer output and the input interface of the apparatus to be controlled (@control' here includes any form of influence) can be deferred to a subsequent stage.

Once a user has defined the scope of a new CPL language (how many words, their general structure, etc.), the CPL creation process can be done in an automatic way by the automatic selection of appropriate combinations of phones that reduce confusion.

Several levels of CPL can be distinguished according to the degree of sophistication involved:

Level 1 CPL - has a small vocabulary of discrete words that can used for command and control applications. Such a CPL can be used for controlling appliances. For example: a user could use a CPL word to turn the lights on in their home.; a CPL could be created to give true hands-free operation of a phone handset; a universal device-control 'Esperanto' could be used to talk with appliances and computers anywhere; etc.

Level 2 CPL - gives the user enough vocabulary and sentence structure to talk about a particular topic in one domain. For example, the user could talk with a computer-based agent about shipping movements using the CPL language for Shipping.

Level 3 CPL - is a language with a vocabulary size and sentence structure sophistication approaching a traditional Pidgin language such as Creole etc. (but, of course, the CPL language has been designed to be very efficient in terms of talking to computers).

Example

The following example of a CPL is a Level 1 CPL generated with reference to the Figure 1 matrix.

Consider the case of a simple telephone that can be controlled via a set of 16 CV words to perform the following actions.

Switch on
Switch off
Redial
Dial <number>
Hangup
Answer
<number> = 0,...9.

By looking at the Figure 1 matrix, it is possible to find some phones that fit well together (i.e minimize the confusion) to produce these 16 words.

Switch on:	[g ow]	as in go
Switch off:	[s ow]	as in so
Answer	[h aa l uw]	as in ha loo
Hangup	[b aa]	as in baa
Dial	[d ao d ao]	as in dodo (o of born)
Redial	[r aa d ao d ao}	as in raadodo.
Zero	[p ow p ow]	as in popo
One	[n ow n ow}	as in nono
Two	[t aa t aa]	as in tata
Three	[g ao g ao]	as in gogo
Four	[f ow f ow]	as in fofo

Five	[f aa f aa]	as in fafa
Six	[s aa s aa]	as in sasa
Seven	[s uw s uw]	as in soosoo
Height	[h ow h ow]	as in hoho
Nine	[n aa n aa]	as in nana

Usage

A CPL, once created, would be used in much the same way as an existing language subset to inform a speech recogniser of the lexicon and grammar of the language being recognised. Since many different CPLs can be defined, then, at least in device control applications where the number of words in each CPL is likely to be limited, it is convenient to enable a speech recogniser to be capable of recognising multiple CPLs, though the recogniser is preferably informed of the identity of the CPL being used for a current voice input. Thus, with reference to Figure 2, three voice-controlled devices 30, each with voice-input microphone M but each having a different associated CPL, use the same speech recogniser 32. Whenever a device 30 wishes to use the recogniser, it first sends its identity to a CPL select block of the recogniser which looks up (in a table held memory) the CPL appropriate for the device, retrieves this CPL from store 34, and loads the CPL into the recogniser. The recogniser is now set up to interpret the CPL received form the device concerned with the recogniser output being fed back to that device.

Training

Figure 4 illustrates a system for training a user to use a selected CPL. The training is controlled by a training manager function 40 (for example, an application running on a PC). The training manager, after having been told by the user for which CPL language training is required, communicates with a text-to-speech (TTS) converter 40 to run training scripts held in store 42; the manager 40 can also generate text-based messages itself for interpretation by the TTS converter 41. The output of converter 41 is fed to the user 2 over loudspeaker 43.

The user 2 responds to the selected training script (for example, by repeating a converter-spoken CPL word or phrase) and the user's spoken response is picked up by microphone 44 and passed to a speech analysis and recognition block 45 that is informed by the lexicon and grammar 46 of the selected CPL. The output of the speech recogniser 45 and the output expect by the training script are both passed to error detector block 47 that detects and analyses errors in the user's input. The error information is passed to the training manager function 40 that determines what action to take next (repetition of a mis-pronounced word, etc).

Variants

It will be appreciated that many variants are possible to the above described embodiments of the invention.

CLAIMS

1. Apparatus provided with a voice-input interface including a speech recogniser adapted to recognise input in the form of an artificial language made up of words formed from a set of phonemes, or other utterance elements, chosen for their low risk of confusion one with another by the speech recogniser.

2. A method of controlling apparatus wherein a user uses an artificial language to instruct the apparatus by voice input, the apparatus having an associated speech recogniser and said artificial language being made up of words formed from a set of phonemes, or other utterance elements, chosen for their low risk of confusion one with another by the speech recogniser.

3. An artificial spoken language for apparatus control, the language being made up of words formed from a set of phonemes, or other utterance elements, chosen for their low risk of confusion one with another by a speech recogniser.

4. A training system for teaching a person to use an artificial language, the system including means for receiving voice input from the user, means for detecting errors in the pronunciation and/or grammar of the voice input as compared to an expected input that conforms with said artificial language, and means for providing feedback to the user concerning the detected errors, said artificial language being made up of words formed from a set of phonemes, or other utterance elements, chosen for their low risk of confusion one with another by a speech recogniser.

5. Apparatus for use in constructing an artificial language, the apparatus comprising means for providing a set of phonemes or other utterance elements chosen for their low risk of confusion one with another by a speech recogniser, user input means for creating words by assembling together said phonemes/utterance elements, and means for storing the set of created words as a new said artificial language.

6. The claimed subject of any one of the preceding claims, wherein the language includes associated grammar rules.

7. The claimed subject of any one of the preceding claims, wherein said phonemes / utterance elements are all chosen from the same human language.

8. The claimed subject of claim 7, wherein the human language is Esperanto or English..

9. A speech recogniser for use by multiple devices that have different associated artificial languages each according to claim 3, the speech recogniser being adapted to receive in association with particular voice input, an indication of the identity of the artificial language used for the input.

ABSTRACT

Speech Recognition

New spoken languages are provided that can be easily understood by automated speech recognizers associated with equipment, the languagse being learnt by human users in order to speak to the equipment. These new languages are simplified in terms of vocabulary and structure and are specifically designed to minimize recognition errors by automated speech recognizers by being made up of phonemes or other uttered elements that are not easily confused with each other by a speech recognizer. The uttered elements are preferably chosen from an existing language. Apparatus and methods for controlling equipment using these recognizer-friendly languages are also provided as are training systems for training human users to speak these languages, and methods and systems for creating new language instances.

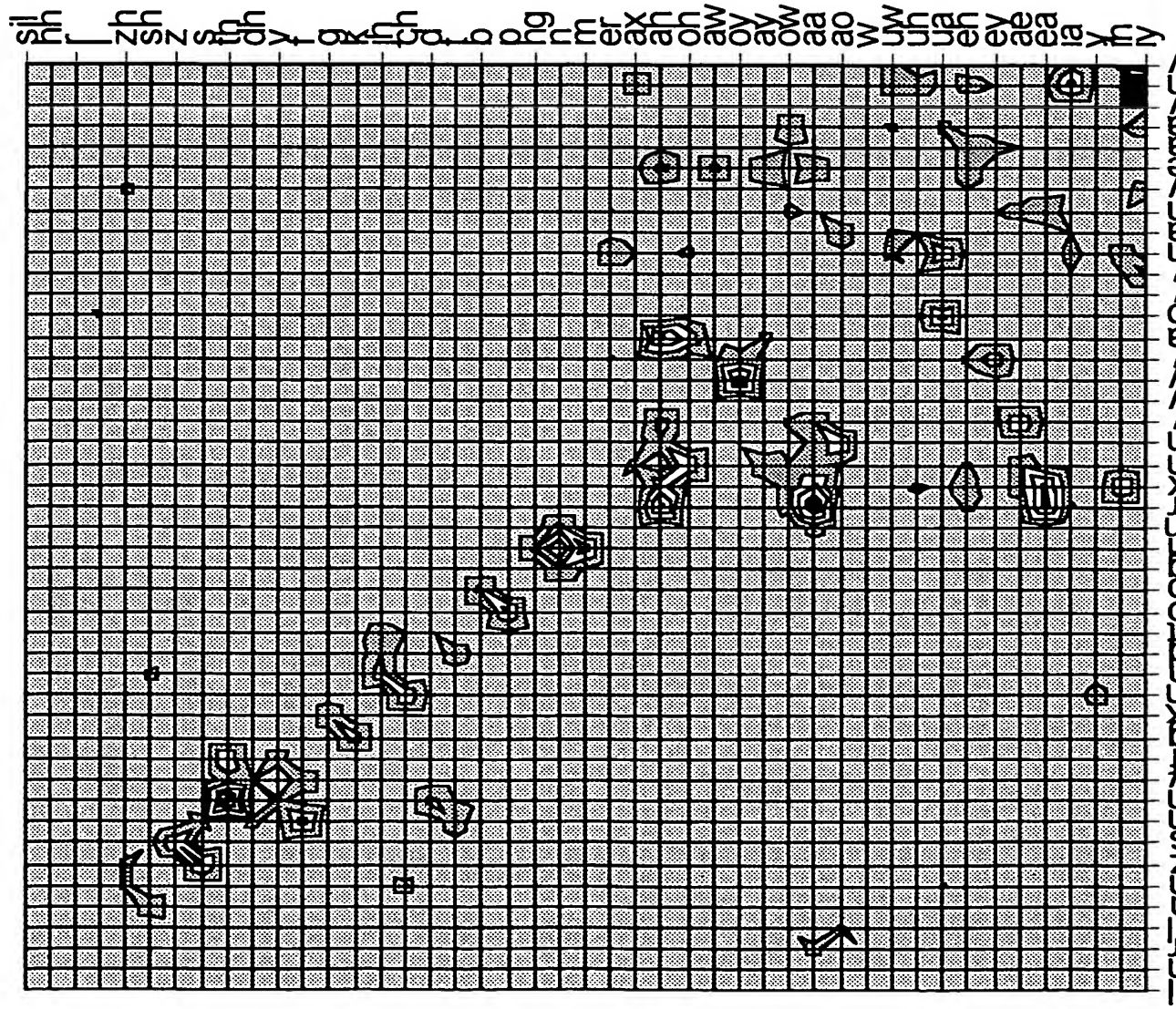
14.00-16.00
12.00-14.00
10.00-12.00
8.00-10.00
6.00-8.00
4.00-6.00
2.00-4.00
0.00-2.00

Figure 1

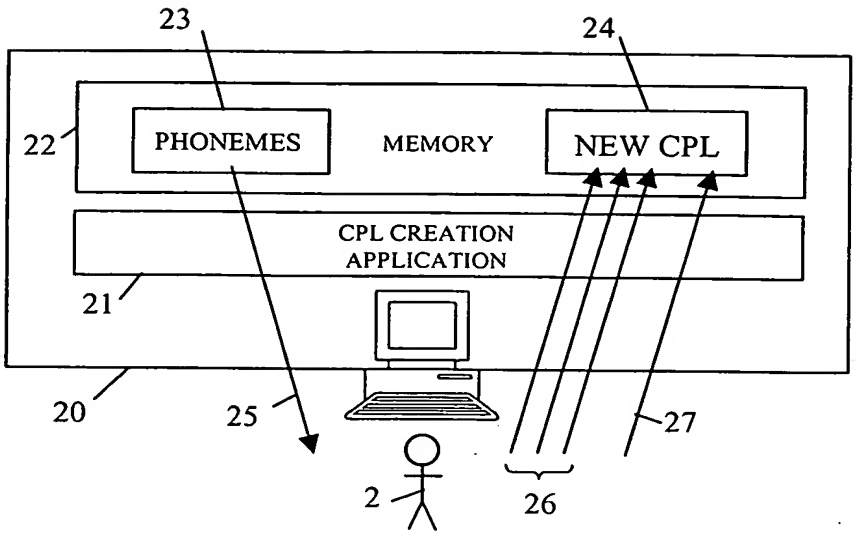

Figure 2

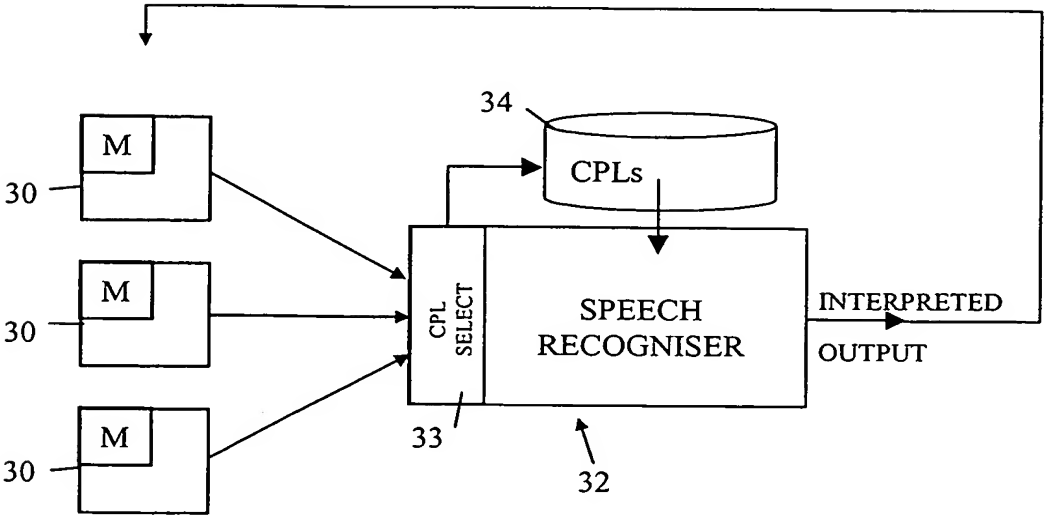

Figure 3

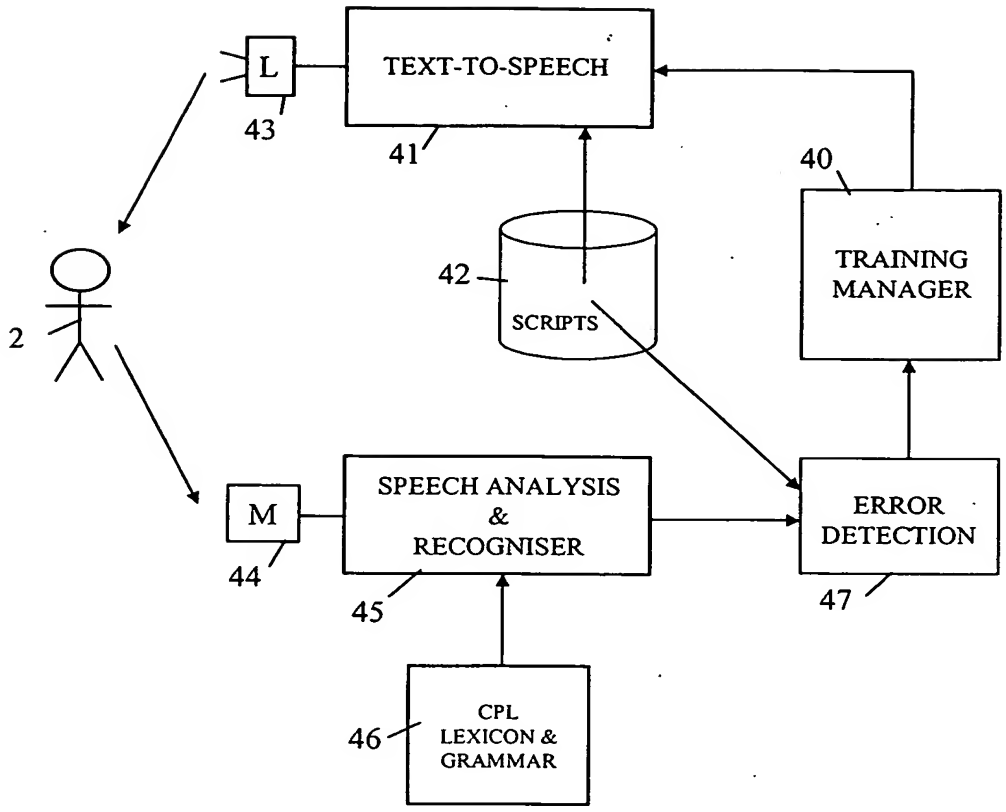
L
43
TEXT-TO-SPEECH
41
40
TRAINING MANAGER
42
SCRIPTS
2
M
44
SPEECH ANALYSIS & RECOGNISER
45
ERROR DETECTION
47
CPL LEXICON & GRAMMAR
46

Figure 4